INTRINSIC DISORDER AND PROTEIN EVOLUTION:

AMINO ACID COMPOSITION OF PROTEINS IN LAST UNIVERSAL ANCESTOR.

Sai Harish Babu Karne

Submitted to the faculty of the School of Informatics

in partial fulfillment of the requirements

for the degree of

Master of Science in Bioinformatics

Indiana University

December 2008

Accepted by the Faculty of Indiana University,

In partial fulfillment of the requirements for the degree of Master of Science

in Bioinformatics

Master's Thesis Committee

_____

Dr. Pedro Romero, PhD, Chair.

_____

Dr. Keith A Dunker, PhD.

_____

Dr. Vladimir N. Uversky, PhD.

Dedicated to My Parents.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

The graduate education at School of Informatics at Indiana University – Purdue University – Indianapolis has been an excellent experience. Now I would like to take the opportunity to express my gratitude to all those who made my graduate studies a memorable experience and made this thesis possible.

I would like to thank my advisor Dr. Pedro Romero for his continuous guidance and support during the course of my thesis. His constant feedback and encouragement helped me lay a strong foundation for the thesis. I would also thank him for his encouragement in learning and exploring new concepts.

I would also like to thank Dr. Vladimir N Uversky and Dr. Keith A Dunker for being a part of my advisory committee and providing valuable inputs towards the thesis. Their meticulous efforts to review my work are greatly appreciated. Also, I would like to add my thanks to the faculty and staff of the Department of Bioinformatics for their cooperation.

I would like to thank my family for their unstinting support and for providing an opportunity to study at IUPUI. Last but not the least; I would like to thank all my teammates of the project, Christopher J Oldfield and Wai Chan for their cooperation and all my friends, Pradeep, Premchand, Vamshi, Amar, Srikant and Rahul for their encouragement and assistance throughout the course of my study.

ABSTRACT

SAI HARISH BABU KARNE

INTRINSIC DISORDER AND PROTEIN EVOLUTION: AMINO ACID COMPOSITION OF PROTEINS IN LAST UNIVERSAL ANCESTOR

All twenty amino acids did not appear simultaneously in nature. Instead some of them appeared early, while others were added into the genetic code later. The amino acids that were formed by Miller (1953) are suggested to have appeared early in evolutionary history, and the amino acids associated with codon capture developed late in the course of evolution. The chronological order of appearance of the amino acids proposed by Trifonov (2000) was G/A, V/D, P, S, E/L, T, R, N, K, Q, I, C, H, F, M, Y, W. According to Romero et al. (1997) amino acids G, D, E, P and S are disorder-promoting residues and C, F, W and Y are order-promoting residues this means that the early or the ancient amino acids were disorder promoting and the order promoting residues came late into the genetic code. These observations led to the hypothesis that the first proteins, which were comprised of the early amino acids only, were disordered, and, furthermore, that the appearance of the late amino acids and the appearance of the structural proteins were concurrent. Software developed by Brooks et al. (2004) to find the amino acid composition of the LUA (Last Universal Ancestor) was used to test this hypothesis. For this work, the Clusters of Orhtologous Groups of proteins (65 COGs) were split into enzymes and non-enzymes. It was found that intrinsic disorder was abundant in both the groups of proteins, with non enzymes being much more disorder than enzymes. Further analysis was done to check for the frequency of the modern amino acids C, F, W, and Y in the Protein data bank (PDB) and Swissprot.